


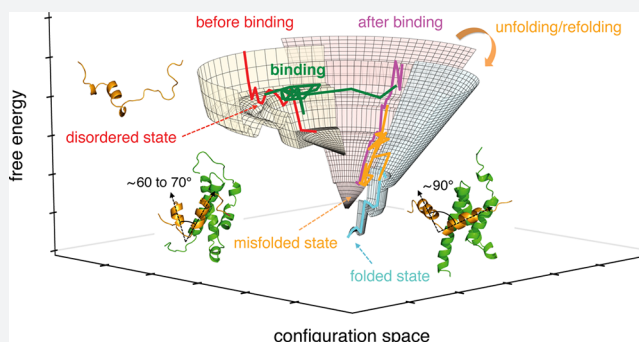
Explicit Characterization of the Free Energy Landscape of pKID–KIX Coupled Folding and Binding

Song-Ho Chong, Haeri Im, and Sihyun Ham*

Department of Chemistry, The Research Institute of Natural Sciences, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 04310, Korea

Supporting Information

ABSTRACT: The most fundamental aspect of the free energy landscape of proteins is that it is globally funneled such that protein folding is energetically biased. Then, what are the distinctive characteristics of the landscape of intrinsically disordered proteins, apparently lacking such energetic bias, that nevertheless fold upon binding? Here, we address this fundamental issue through the explicit characterization of the free energy landscape of the paradigmatic pKID–KIX system (pKID, phosphorylated kinase-inducible domain; KIX, kinase interacting domain). This is done based on unguided, fully atomistic, explicit-water molecular dynamics simulations with an aggregated simulation time of $>30 \mu\text{s}$ and on the computation of the free energy that defines the landscape. We find that, while the landscape of pKID before binding is considerably shallower than the one for a protein that autonomously folds, it becomes progressively more funneled as the binding of pKID with KIX proceeds. This explains why pKID is disordered in a free state, and the binding of pKID with KIX is a prerequisite for pKID's folding. In addition, we observe that the key event in completing the pKID–KIX coupled folding and binding is the directed self-assembly where pKID is docked upon the KIX surface to maximize the surface electrostatic complementarity, which, in turn, require pKID to adopt the correct folded structure. This key process shows up as the free energy barrier in the pKID landscape separating the intermediate nonspecific complex state and the specific complex state. The present work not only provides a detailed molecular picture of the coupled folding and binding of pKID but also expands the funneled landscape perspective to intrinsically disordered proteins.



■ INTRODUCTION

The free energy landscape underlies the thermodynamics and kinetics of any molecular processes in solution. It is the graph of the free energy (to be denoted as f) across the configuration space whose point (collectively abbreviated as \mathbf{r}) is specified by the coordinates of all atoms constituting a molecule of interest.^{1–3} (Figure 1 illustrates three types of “free energies” used in different contexts in the literature to discriminate f from the others.) The free energy f is a sum of the potential energy and the solvation free energy,^{1–3} and the landscape defined by f is a natural extension of the potential energy landscape or surface to systems in which solvation effects cannot be neglected. f determines thermodynamic properties of the system through the partition function, $Z = \int d\mathbf{r} e^{-\beta f(\mathbf{r})}$, in which β denotes the inverse temperature.^{1–3} It also dictates kinetic properties as the force $-\partial f(\mathbf{r})/\partial \mathbf{r}_i$ under the influence of the frictional and thermal random forces, governs the equation of motion for each (say, i th) constituent atom.² The concept of a free energy landscape has, in particular, played a key role in advancing our understanding of protein folding. For example, the basic assumption behind the well-known paradox by Levinthal⁴ is that all possible protein configurations are

equally probable, which amounts to assuming constant $f(\mathbf{r})$, i.e., a flat free energy landscape. It was a keen insight that the paradox is resolved if the free energy landscape is not flat but is globally funneled, i.e., if there is an overall negative slope ($\partial f(\mathbf{r})/\partial \mathbf{r} < 0$) toward the folded state.^{5,6} The funneled landscape perspective has since then served as a conceptual framework not only for interpreting folding experiments^{7,8} but also for understanding a variety of processes, including biomolecular recognition and protein misfolding and aggregation.^{9–12}

New attempts have emerged in recent years that aim to go beyond the conceptual level and to perform quantification of the free energy landscape.^{13,14} However, this is challenging not only because it requires sampling relevant configurations associated with the process of interest but also because efficiently computing the free energy f for all of those individual configurations is nontrivial. Recent advances in computing power and simulation software have opened up the possibility of providing fully atomistic long-time simulation

Received: February 28, 2019

Published: July 15, 2019

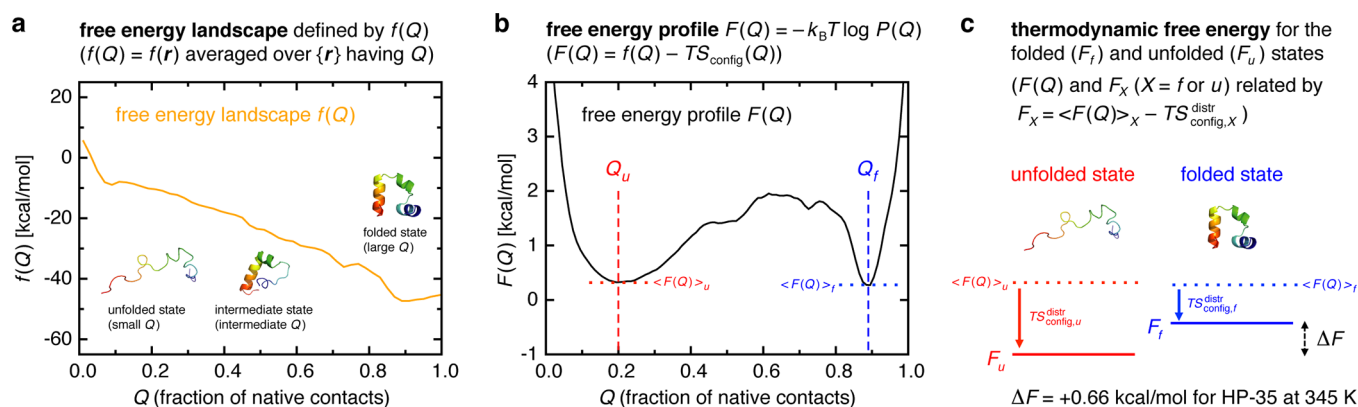


Figure 1. Illustration of the free energy landscape defined by $f(Q)$ (a), the free energy profile $F(Q)$ (b), and the thermodynamic free energy for the folded (F_f) and unfolded state (F_u) (c) based on the $\sim 400 \mu\text{s}$ folding–unfolding simulation trajectory of HP-35. The free energy landscape $f(Q)$ was constructed after computing $f(\mathbf{r})$ for $\sim 2 \times 10^6$ individual configurations saved with a 200 ps interval. The free energy profile $F(Q)$ was obtained from a histogram of Q values sampled in the simulation. The dashed vertical lines in panel b indicate the positions of $Q_u = 0.20$ and $Q_f = 0.89$ that characterize the unfolded-state ($Q < Q_u$) and folded-state ($Q > Q_f$) regions. The folding free energy $\Delta F = F_f - F_u$ was estimated from the population ratio of the folded ($Q > Q_f$) and unfolded ($Q < Q_u$) configurations.

trajectories of protein folding and protein–protein association.^{15–17} An analysis method has also been developed for evaluating thermodynamic functions (including f) along protein configurational changes, which has been utilized to derive thermodynamic perspective on protein aggregation.¹⁸ Combining these developments, it is now possible to construct the free energy landscape from first-principles (Figure 1a provides such an example).

Coupled folding and binding is an intriguing process in which an intrinsically disordered protein (IDP) folds into an ordered structure upon binding with a partner protein.^{19–21} Such conformational flexibilities are central to the functions of numerous IDPs, enabling them to serve as molecular switches and hubs in biological networks.^{22,23} A number of questions naturally arise concerning the fundamental aspects of this process, such as the following: (1) What distinguishes an IDP from proteins that autonomously fold? (2) How are those distinguishing characteristics of an IDP altered, when its binding partner approaches nearby, so that its folding becomes possible? (3) Does the folding occur before, after, or concomitantly with binding? It will be illuminating in this regard to explicitly characterize the free energy landscape of coupled folding and binding as the molecular mechanisms that address these questions are expected to manifest themselves as its topographic characteristics. In addition, it is of significant interest to explore to what extent the funneled landscape perspective applies also to intrinsically disordered proteins.

Here, we investigate the free energy landscape of the phosphorylated kinase-inducible domain (pKID; residues 116–149) of the CREB protein in the presence of its binding partner, the kinase interacting domain (KIX; residues 586–672) of the CREB binding protein. This is a paradigmatic system that exhibits coupled folding and binding—disordered pKID folds into a structure possessing two α helices (termed α_A and α_B) upon binding to KIX (Figure 2a)—and has been intensively investigated through experimental and computational studies.^{24–42} We carry out spontaneous binding simulation starting from disordered pKID placed at a large separation from KIX, to sample the configurations during the course of the coupled folding and binding. To the best of our knowledge, this is the first unguided, fully atomistic pKID–KIX binding simulation that reaches the specific pKID–KIX

complex structure. For each of the simulated configurations (\mathbf{r}), we compute $f(\mathbf{r})$ by applying the molecular integral-equation theory for solvation free energy and then construct the free energy landscape. While rigorous calculation methods for $f(\mathbf{r})$ are available such as free energy simulations, the use of an approximate solvation theory is inevitable here since $f(\mathbf{r})$ for millions of simulated configurations needs to be computed. We analyze in detail how the topography of the free energy landscape varies as pKID binds to KIX. Supplemented with structural analysis of the simulation trajectory, we would like to provide a detailed molecular picture of the coupled folding and binding of the pKID–KIX system.

RESULTS

Free Energy That Defines the Free Energy Landscape. Before embarking on an analysis of pKID–KIX binding, we would like to elaborate on the free energy $f(\mathbf{r})$ that defines the free energy landscape. We feel this is necessary as $F(Q) = -k_B T \log P(Q)$, another free energy also referred to as the potential-of-mean-force curve, which is associated with the probability distribution $P(Q)$ of a certain order parameter (or reaction coordinate) Q , has often been identified in the literature as the “free energy landscape”; however, its definition is distinct from the definition introduced in refs 5 and 6 for discussing the funneled landscape. It is therefore instructive to refer to the relationship between $f(\mathbf{r})$, $F(Q)$, and the conventional thermodynamic free energy F . In Figure 1, we illustrate these free energies, which were computed on the basis of the $\sim 400 \mu\text{s}$ folding–unfolding simulation trajectory of villin headpiece subdomain (HP-35) provided by the D. E. Shaw Research.⁴³ HP-35 was chosen as it is a 35-residue α -helical protein that autonomously folds⁴⁴ and, hence, serves as a good reference system in discussing the free energy landscape of the 34-residue pKID.

The function $f(\mathbf{r})$ emerges in an attempt to express the partition function Z of the system solely in terms of the configuration \mathbf{r} of a solute molecule of interest (a protein here), $Z = \int d\mathbf{r} e^{-\beta f(\mathbf{r})}$, by integrating out all the degrees of freedom associated with the rest (regarded as solvent) of the system (see Methods in the SI). $f(\mathbf{r}) = E_u(\mathbf{r}) + G_{\text{solv}}(\mathbf{r})$ is given by a sum of the solute potential energy (E_u) and the solvation free energy (G_{solv}). As G_{solv} is involved, it is more appropriate

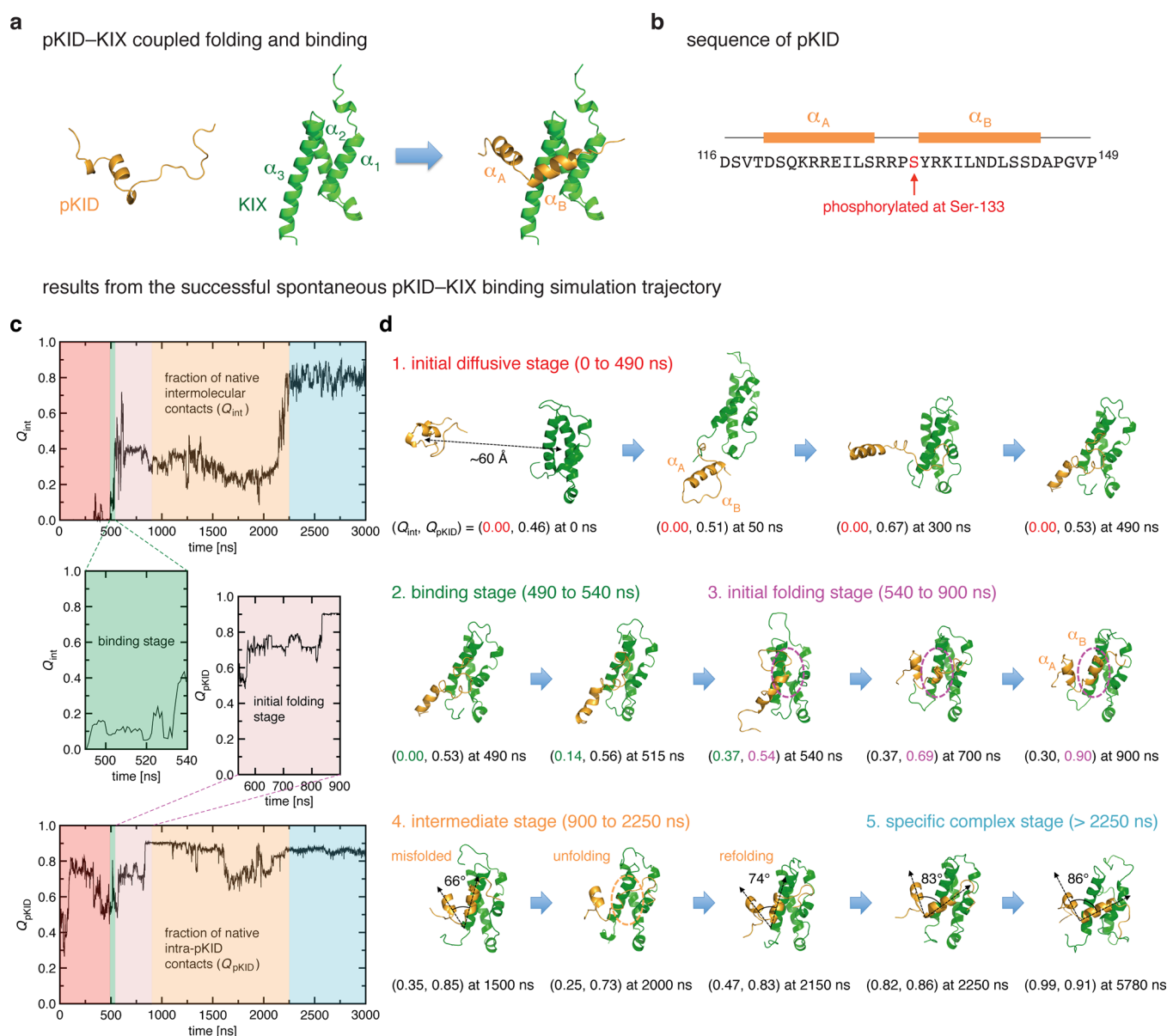


Figure 2. (a) Folding of the disordered pKID upon binding to KIX. The structure of the free pKID is taken from the simulation and those of the free KIX and the pKID–KIX complex from the NMR complex structure (PDB entry 2LXT). (b) Sequence of the pKID region and the locations of the two α helices (α_A and α_B) and of the phosphorylation site. (c) Fraction of native intermolecular contacts Q_{int} (top panel) and fraction of native intra-pKID contacts Q_{pKID} (bottom panel) from the successful 10 μs spontaneous binding trajectory (only the results up to 3 μs are displayed here to facilitate the visibility; the whole results are shown in Figure S3). The time regions are colored as follows: the initial diffusive stage (0–490 ns) colored red; the binding stage (490–540 ns) colored green; the initial folding stage (540–900 ns) colored magenta; the intermediate stage (900–2250 ns) colored orange; and the final specific complex stage (>2250 ns) colored cyan. Q_{int} in the binding stage and Q_{pKID} in the initial folding stage are magnified in the middle panels. (d) Selected structures along with $(Q_{\text{int}}, Q_{\text{pKID}})$ values and the angle between the pKID α_A and α_B helices.

to refer to $f(\mathbf{r})$ as the free energy rather than the potential energy. It is essential to recognize here that $f(\mathbf{r})$ is defined for a single individual configuration \mathbf{r} . As such, it carries no solute configurational entropy. $P(\mathbf{r}) = e^{-\beta f(\mathbf{r})}/Z$ can be interpreted as the probability of observing a specific configuration \mathbf{r} . Therefore, $f(\mathbf{r})$ is directly connected to Levinthal's argument: in fact, the assumption that all possible protein configurations are equally probable amounts to assuming a constant $f(\mathbf{r})$. Correspondingly, $f(\mathbf{r})$ is precisely the quantity with which the funneling concept has been introduced.^{5,6} Here and in the following, we shall use $f(Q)$, which is an arithmetic average of $f(\mathbf{r})$ over the configurations conforming to $Q = Q(\mathbf{r})$, in drawing the free energy landscape as $f(\mathbf{r})$ is defined on a high-

dimensional configuration space and is not suitable for visualization. We observe from Figure 1a that the free energy landscape $f(Q)$ indeed exhibits the funneled character as predicted by the landscape theory: the overall negative slope ($df(Q)/dQ < 0$) toward the folded state serves as a proxy of the genuine funneledness ($\partial f(\mathbf{r})/\partial \mathbf{r} < 0$).

The free energy $F(Q) = -k_B T \log P(Q)$ is associated with the restricted partition function, $P(Q) = Z(Q)/Z$ with $Z(Q) = \int_{Q=Q(\mathbf{r})} d\mathbf{r} e^{-\beta f(\mathbf{r})}$, in which the integration is over the configurations satisfying $Q = Q(\mathbf{r})$. As a number of solute configurations contribute to $F(Q)$, it carries the solute configurational entropy, to be denoted as $S_{\text{config}}(Q)$. Indeed, $f(\mathbf{r})$ and $F(Q)$ are related via $F(Q) = f(Q) - TS_{\text{config}}(Q)$.^{5,45}

The free energy profile $F(Q)$ (Figure 1b) exhibits quite a dissimilar appearance from the free energy landscape $f(Q)$ (Figure 1a). In particular, the transition-state barrier shows up in $F(Q)$. Given the relation $F(Q) = f(Q) - TS_{\text{config}}(Q)$ and the funneled character of $f(Q)$, the barrier has a purely entropic origin.^{46,47} This is one of the major differences between protein folding and simple chemical reactions of small molecules:⁵ in the latter, the barrier appears already in $f(Q)$ as it stems from the high potential energy of a fairly well-defined transition-state structure.

$F(Q)$ can be connected to the thermodynamic free energy F_X by recognizing that the folded ($X = f$) and unfolded ($X = u$) states can be characterized by the respective regions, $Q > Q_f$ and $Q < Q_u$,⁴⁸ along the reaction coordinate. One can then show that $F_X = \langle F(Q) \rangle_X - TS_{\text{config},X}^{\text{distr}}$. Here, $\langle F(Q) \rangle_X$ refers to an average over the Q region associated with state X , and $S_{\text{config},X}^{\text{distr}}$ is the configurational entropy originating from the presence of the distribution of Q values in state X . This explains why the difference between the folded- and unfolded-state minima of $F(Q)$ (indicated by $\langle F(Q) \rangle_f$ and $\langle F(Q) \rangle_u$ in Figure 1b) does not correspond to the thermodynamic folding free energy $\Delta F = F_f - F_u$. In fact, an inspection of Figure 1b indicates that the folding is apparently slightly exothermic ($\langle F(Q) \rangle_f < \langle F(Q) \rangle_u$), which however contradicts the estimated $\Delta F = +0.66$ kcal/mol. This is because the width of $F(Q)$ also contributes to the thermodynamic configurational entropy: the wider width of the unfolded-state region of $F(Q)$ than the one for the folded-state region results in larger $S_{\text{config}}^{\text{distr}}$ for the former, leading to $F_f > F_u$ as shown in Figure 1c.

The free energies $f(\mathbf{r})$, $F(Q)$, and F constitute a natural series ranging from microscopic (defined for individual configurations \mathbf{r}) to intermediate (defined for a set of configurations \mathbf{r} satisfying $Q = Q(\mathbf{r})$) to macroscopic (defined for a state characterized by a range of Q). The equilibration with respect to the solute configurations is therefore essential for a reliable determination of $F(Q)$ and F . On the other hand, this does not apply to $f(\mathbf{r})$ (and $f(Q)$ therefrom) as it is defined for individual solute configurations; i.e., $f(\mathbf{r})$, rather than being an equilibrium property, describes the topography of the configuration space. Indeed, quite a similar free energy landscape for HP-35 can be gained just based on an unequilibrated, single folding portion of the simulation trajectory (Figure S1). This holds even though individual folding pathways of HP-35 are quite heterogeneous.⁴³

Furthermore, since equilibration with respect to solute configurations is not involved, $f(\mathbf{r})$ is more relevant to nonequilibrium dynamical processes. Indeed, while $f(\mathbf{r})$ was introduced above through a thermodynamic argument, it also shows up in the dynamics equation, $m_i \ddot{\mathbf{r}}_i = -\partial f(\mathbf{r}) / \partial \mathbf{r}_i - \gamma^0 \dot{\mathbf{r}}_i + \mathbf{R}_i$, which can be derived by eliminating the solvent degrees of freedom from Newton's equations of motion through the use of a projection-operator formalism (see Methods in the SI for an outline of the derivation). Here, the bare friction coefficient γ^0 and the random force \mathbf{R}_i are connected by the fluctuation–dissipation theorem. This equation clearly indicates that the solute dynamics is dictated by $f(\mathbf{r})$. It also ensures that, after long times, the solute configuration \mathbf{r} is populated proportional to $e^{-\beta f(\mathbf{r})}$.

Structural Changes along the Coupled Folding and Binding. We carried out spontaneous pKID–KIX binding simulations at 300 K and 1 bar. The initial disordered pKID

structure was taken from a separate 1 μ s simulation of the free pKID which was performed beforehand (Figure S2): the average α -helical contents of the α_A ($63.4 \pm 22.1\%$) and α_B ($9.0 \pm 14.3\%$) regions from our simulation are in fair agreement with the experimental observations ($\alpha_A \sim 50$ – 60% and $\alpha_B \sim 15\%$) for the free pKID.²⁵ (We notice, in light of these experimental and simulation results for the free pKID, that it is mainly the folding of the α_B helix that occurs when pKID binds to KIX. We also remark here that the angle between the pKID α_A and α_B helices is $\sim 90^\circ$ in the experimental pKID–KIX complex structure.²⁸) The disordered pKID taken from our free pKID simulation and KIX from the NMR complex structure were initially placed with a ~ 60 Å center-of-mass distance, and we applied no artificial attraction force between them during our simulations. In total, 10 independent binding simulations with at least a length of 1.5 μ s were carried out: two of them were extended to 3 μ s, and one of them was extended to 10 μ s. (All the simulations reported in the present work are summarized in Table S1.) The simulation with a length of 10 μ s resulted in the successful folding of pKID upon binding (the minimum C_α root-mean-square-deviation of the folded pKID from the corresponding NMR structure is 1.0 Å), and we start from analyzing this successful trajectory.

The binding of pKID with KIX and the folding of pKID are monitored via the fraction of native intermolecular contacts (Q_{int}) and the fraction of native intra-pKID contacts (Q_{pKID}), respectively (Figure 2c). Selected structures along the trajectory are also displayed (Figure 2d). In the initial diffusive stage (from 0 to 490 ns; the time region colored red), pKID diffuses and approaches KIX and then explores the surface of KIX with disordered structures ($Q_{\text{pKID}} \sim 0.5$). Certain intermolecular contacts are formed as can be inferred from the displayed structures, which however are mostly non-native ones, and Q_{int} in this stage remains close to 0. The native intermolecular contacts start to be formed from 490 ns by the anchoring of the disordered pKID α_B region onto the KIX surface, and Q_{int} increases to ~ 0.4 by 540 ns (the binding stage; the time region colored green). Then, we observe the folding of the pKID α_B helix (dashed magenta ellipses in Figure 2d), and Q_{pKID} increases from ~ 0.5 to ~ 0.9 in the initial folding stage (from 540 to 900 ns; the time region colored magenta). However, the folding of pKID is still incomplete in that the angle between the α_A and α_B helices ($\sim 60^\circ$ to 70°) significantly deviates from the one ($\sim 90^\circ$) found in the experimental folded structure, and such pKID structures shall be referred to as the “misfolded” structures. Correspondingly, the native intermolecular contacts are not fully formed between the misfolded pKID and KIX ($Q_{\text{int}} \sim 0.3$ – 0.4). We term the stage characterized by the misfolded pKID and deficient intermolecular contacts as the intermediate stage (from 900 to 2250 ns; the time region colored orange). This stage is terminated by the unfolding of the α_B helix (dashed orange ellipse in Figure 2d), which is followed by a refolding that accompanies the widening of the α_A – α_B angle toward $\sim 90^\circ$. Such conformational changes are more clearly reflected in the pKID α_B helix content and the α_A – α_B angle along the trajectory presented in Figure S4. Finally, the system enters the stage where the pKID–KIX complex takes structures close to the experimental one (>2250 ns; the time region colored cyan).

Free Energy Landscape of the Coupled Folding and Binding. The presence of various distinctive stages during the

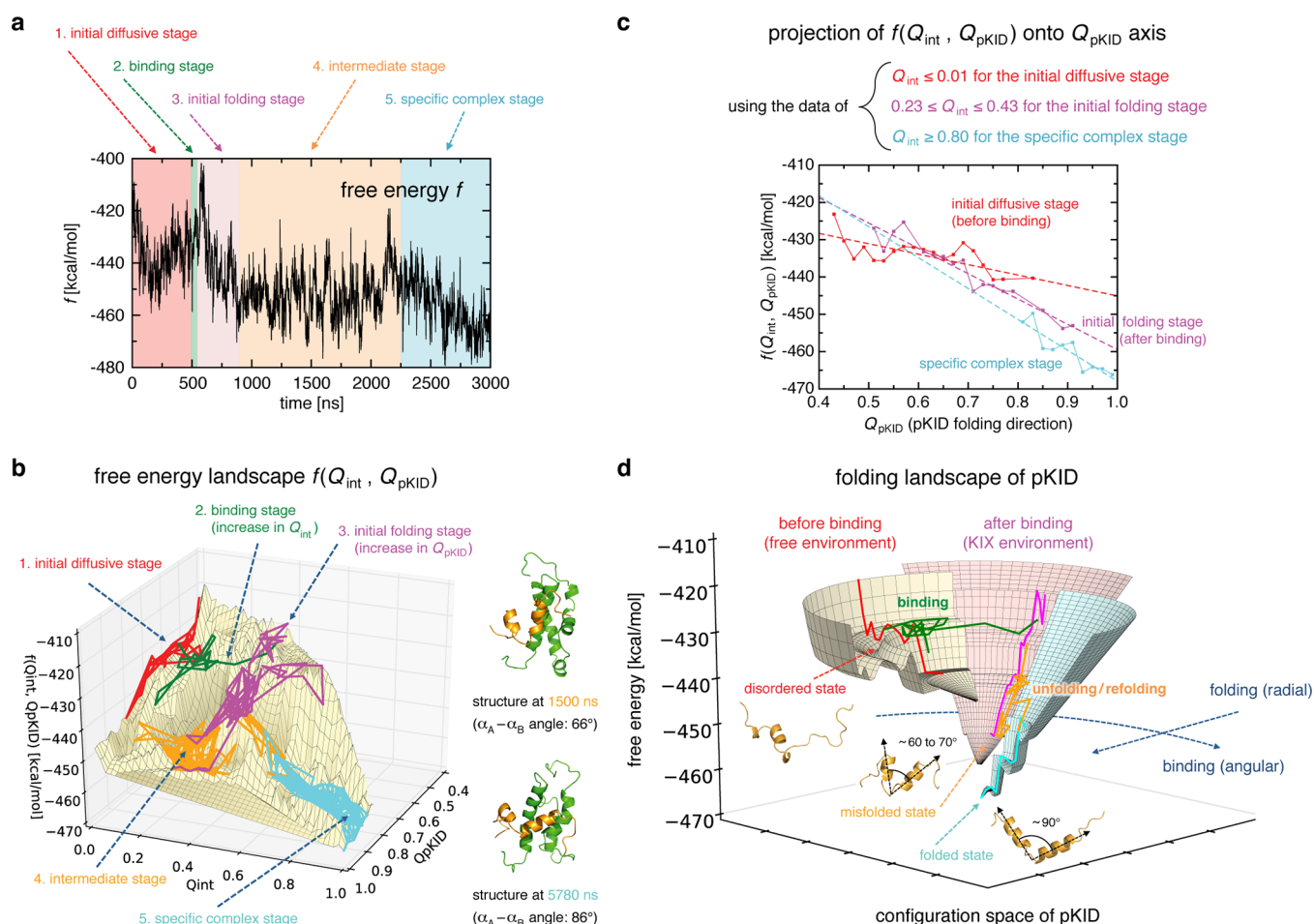


Figure 3. (a) Free energy $f = f_{\text{pKID}} + \Delta f_{\text{int}}$ versus the simulation time from the successful 10 μs spontaneous pKID–KIX binding trajectory. The time regions are colored as in Figure 2c to highlight the various stages of the binding process (only the result up to 3 μs is displayed here; the whole result is presented in Figure S5). (b) Free energy landscape $f(Q_{\text{int}}, Q_{\text{pKID}})$ constructed from the successful binding trajectory. The trajectory projected onto the landscape is drawn with a solid line, whose portions are colored as in panel a. (c) Projection of $f(Q_{\text{int}}, Q_{\text{pKID}})$ onto the Q_{pKID} axis. The dashed lines denote the linear fits to the respective curves of the same color. (d) Folding landscape of pKID constructed from the slopes of the curves shown in panel c (see Figure S6 for details of its construction).

pKID–KIX coupled folding and binding is manifested in the underlying free energy f . We computed f along the successful 10 μs spontaneous pKID–KIX binding trajectory (Figure 3a; only the result up to 3 μs is displayed to facilitate the visibility of various stages, and the result for the entire time range is shown in Figure S5). Here, $f = f_{\text{pKID}} + \Delta f_{\text{int}}$ is a sum of the free energy f_{pKID} for the folding of pKID and $\Delta f_{\text{int}} = \Delta E_{\text{int}} + \Delta G_{\text{solv}}$ associated with the pKID–KIX binding. In the latter, ΔE_{int} is the direct (gas-phase) pKID–KIX interaction potential, and ΔG_{solv} is the solvent-induced one defined by $G_{\text{solv}}(\text{pKID:KIX}) - [G_{\text{solv}}(\text{pKID}) + G_{\text{solv}}(\text{KIX})]$.⁴⁹ Several observations can be made from Figure 3a: Overall, f decreases as the system goes through the initial diffusive (the time region colored red), intermediate (orange), and final (cyan) stages. f decreases when the pKID folding occurs in the initial folding stage (the time region colored magenta), and the free energy barrier is discernible at the end of the intermediate stage (the time region colored orange). However, a more illuminating plot can be gained by combining f (Figure 3a) with Q_{int} and Q_{pKID} (Figure 2c): that is, the free energy landscape $f(Q_{\text{int}}, Q_{\text{pKID}})$ of coupled folding and binding expressed over the binding (Q_{int}) and folding (Q_{pKID}) coordinates.

The free energy landscape constructed from the successful 10 μs pKID–KIX binding simulation is shown in Figure 3b. The landscape $f(Q_{\text{int}}, Q_{\text{pKID}})$ was obtained by dividing the configuration space spanned by $0 \leq Q_{\text{int}} \leq 1$ and $0 \leq Q_{\text{pKID}} \leq 1$ into 2500 small areas after discretizing each of Q_{int} and Q_{pKID} into 50 bins and then computing the average f value from those configurations that belong to each small area. The trajectory projected onto the landscape is drawn with a solid line, whose portions are colored as before to highlight the various distinctive stages. In the initial diffusive stage (colored red), the system exhibits the “vertical” diffusive dynamics along the folding (Q_{pKID}) coordinate that is restricted to the $Q_{\text{int}} \approx 0$ region. This is followed by the binding stage (green) that mostly occurs “horizontally” along the Q_{int} coordinate. Then, the folding of pKID occurs (magenta) nearly vertically along the folding (Q_{pKID}) direction at $Q_{\text{int}} \sim 0.3$ – 0.4 .

What distinguishes the stage before binding, where pKID remains disordered, and the stage after binding, where the folding of pKID occurs? To elucidate this fundamental point, we projected the landscape $f(Q_{\text{int}}, Q_{\text{pKID}})$ onto the Q_{pKID} axis to estimate the landscape’s slopes along the folding (Q_{pKID}) direction (Figure 3c) and then constructed the folding landscape of pKID, shown in Figure 3d, based on those

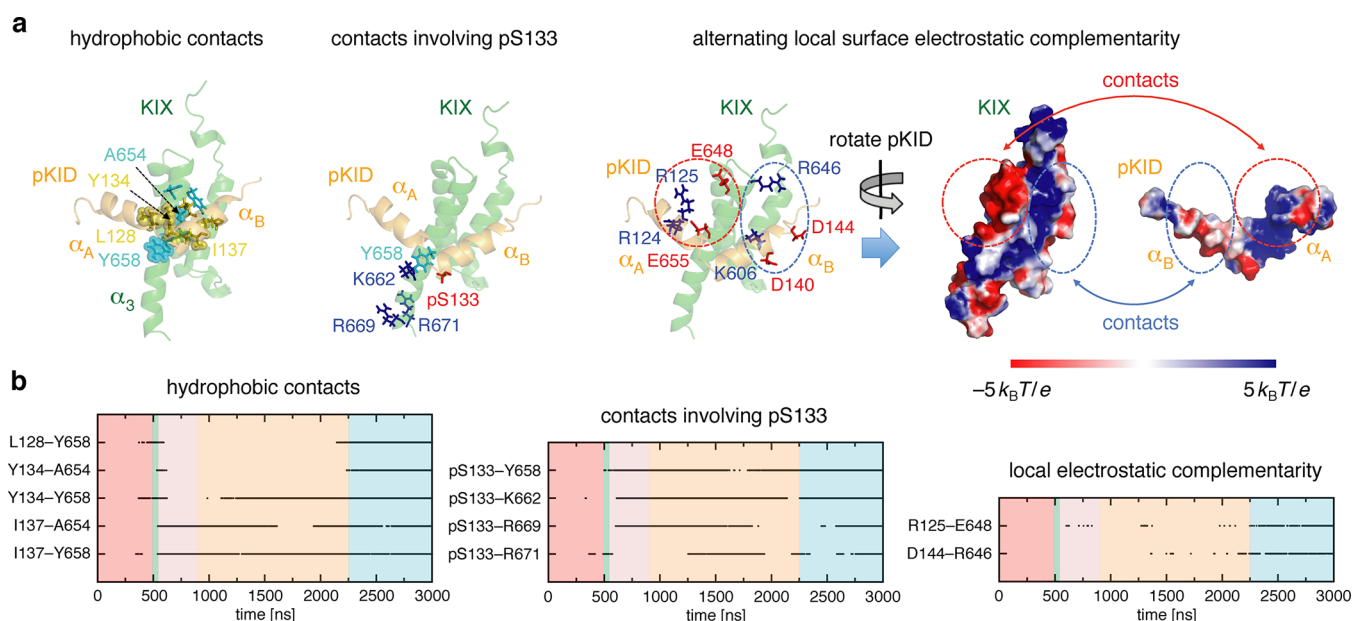


Figure 4. (a) Amino acid residues forming representative intermolecular hydrophobic contacts (left); those forming the intermolecular contacts with the phosphorylated Ser-133 (middle); and those associated with the alternating local surface electrostatic complementarity (right). In the electrostatic potential surface representation, pKID is rotated such that the binding areas are facing the reader. (b) Presence (indicated by a dot) of the intermolecular residue contacts versus the simulation time. The time regions are colored as in Figure 2c (only the results up to 3 μ s are displayed here; the whole results are presented in Figure S9).

slopes. (Our focus in the “funneled” diagram displayed in Figure 3d is the landscape’s slope for which the funnel concept was originally introduced.⁵ In this regard, we notice that, unlike typical funnel diagrams in which the width of the funnel is meant to represent the configurational entropy, the width in Figure 3d is associated with the presence of the binding direction.)

As the slope of the landscape characterizes the landscape’s funneledness, measuring the strength of the energetic bias toward the folded state, Figure 3d clearly indicates (i) that pKID remains disordered before binding as the landscape is not steep enough to allow it to fold, (ii) that the landscape becomes more progressively funneled as the binding proceeds, and (iii) that the folding of pKID occurs when the landscape is steep enough. Indeed, although the free energy landscape of pKID in a free environment is significantly shallower than the one of HP-35 which autonomously folds (we notice in this regard that the pKID landscape before binding agrees well with the one from the free pKID simulation; see Figure S7), the slope of the pKID landscape in the KIX environment becomes comparable to that of the HP-35 landscape as demonstrated in Figure S8.

In the intermediate stage (colored orange in Figure 3b), pKID exists in its misfolded state as mentioned above. The unfolding and refolding of the misfolded pKID that occur at the end of the intermediate stage appear as the barrier region on the landscape preceding the final specific complex stage (colored cyan in Figure 3b). Remarkably, we observe an additional increase in the slope of the landscape in the final stage (Figure 3c,d). This provides further energetic bias to complete the folding of pKID, and the resulting folded state is lower in free energy than the misfolded state.

Native Contact Formation during the Binding of pKID to KIX. So far, we have primarily focused on the folding of pKID, i.e., the folding (Q_{pKID}) direction of the free energy landscape. Here, we investigate the molecular details on the

formation of native intermolecular contacts during the spontaneous binding of pKID to KIX, i.e., the horizontal (Q_{int}) direction of the landscape. Prior to such an investigation, we performed six independent 1 μ s equilibrium pKID–KIX complex simulations to analyze the nature of the native intermolecular contacts. Table S2 lists the representative contacts obtained therefrom. We identified the following three types of native intermolecular contacts (Figure 4a): the hydrophobic contacts between residues that belong to the pKID α_B helix and the KIX α_3 helix; the contacts involving the phosphorylated Ser-133; and the electrostatic contacts that reflect the alternating local electrostatic complementarity between the binding surfaces. The formation of these contacts during the spontaneous binding is shown in Figure 4b.

We find that the successful spontaneous binding is initiated by the contact formation between the hydrophobic residues of pKID and Tyr-658 of KIX (the time region colored red in Figure 4b). This is followed by the formation of the contacts involving the phosphorylated Ser-133 (the time region colored green). In the initial folding and intermediate stages (the time regions colored magenta and orange), the hydrophobic contacts and the contacts involving the phosphorylated Ser-133 are only partially formed, and the electrostatic contacts associated with the local electrostatic complementarity are barely formed. This reflects the fact that the misfolded pKID structures in these stages, in which the pKID α_A – α_B angle is ~ 60 – 70° , are not optimum for making intermolecular contacts with KIX. Indeed, after the unfolding/refolding of the misfolded pKID that occur at the end of the intermediate stage, resulting in the pKID α_A – α_B angle of $\sim 90^\circ$, most of the native intermolecular contacts are formed (the time region colored cyan). In particular, the aforementioned electrostatic contacts are now present. This is because, for those electrostatic contacts to be formed, pKID must be present on the KIX surface not only at the right place but also with the correctly folded structure. Such electrostatic interactions

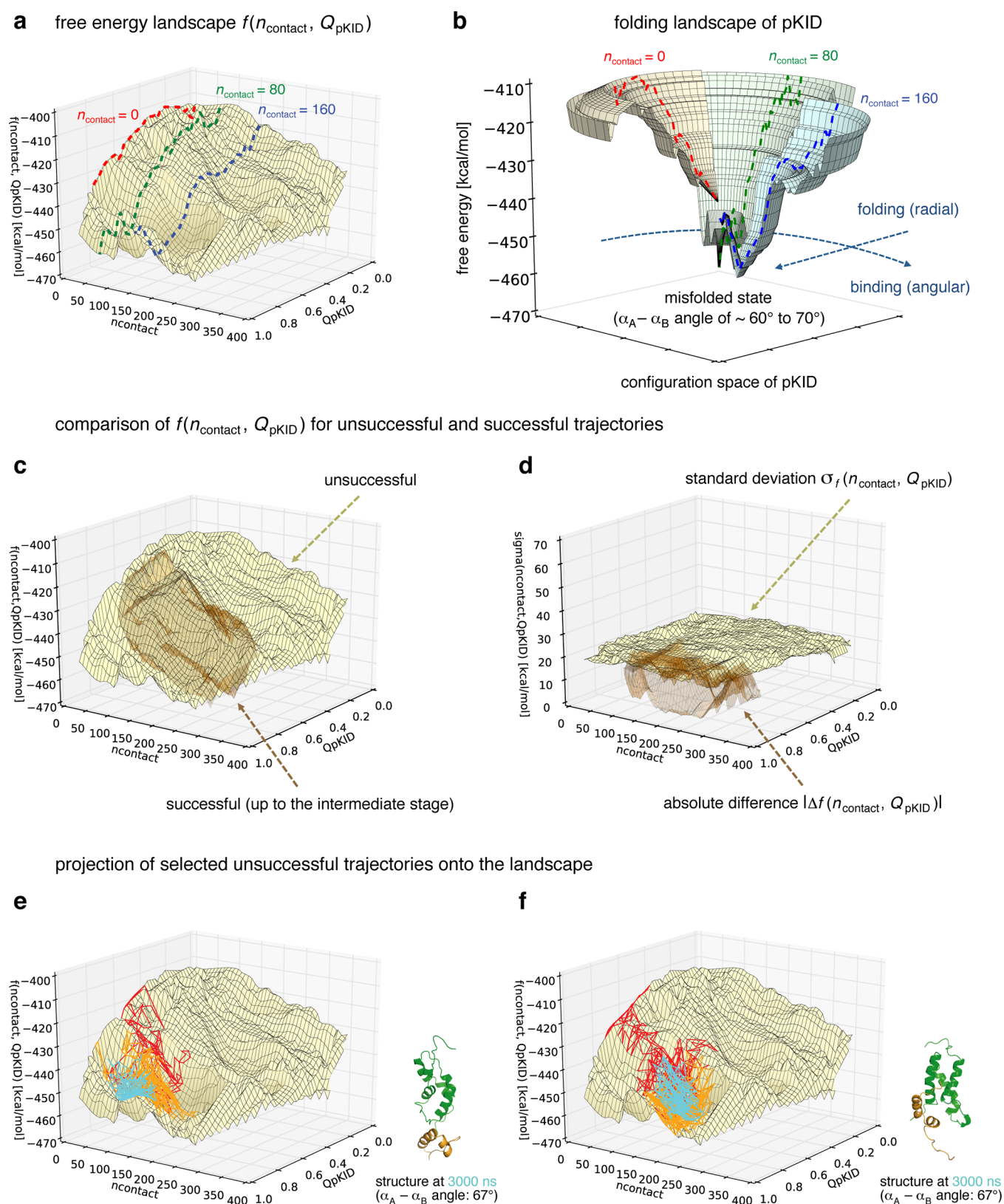


Figure 5. (a) Free energy landscape $f(n_{\text{contact}}, Q_{\text{pKID}})$ constructed from the nine unsuccessful pKID–KIX binding trajectories (surface colored yellow). Sections along Q_{pKID} taken at $n_{\text{contact}} = 0, 80,$ and 160 are shown by red, green, and blue dashed lines, respectively. (b) Folding landscape of pKID constructed from these sections. (c) Comparison of $f(n_{\text{contact}}, Q_{\text{pKID}})$ for the unsuccessful (surface colored yellow) and successful (brown) trajectories. The latter landscape is a redrawing of the one in Figure 3b with the binding coordinate n_{contact} and using only the trajectory data up to the intermediate stage (<2250 ns). (d) Standard deviation computed from the unsuccessful trajectories and the absolute difference between the landscapes for the unsuccessful and successful trajectories. (e) A selected trajectory projected onto the landscape is drawn with a solid line that is colored red, orange, and cyan every 1000 ns. (f) Corresponding result based on another trajectory.

should therefore be crucial for the pKID–KIX binding specificity. Those intermolecular contacts formed only after the unfolding/refolding of the misfolded structure are responsible for the additional increase in the landscape slope discussed above in connection with Figure 3c,d. Indeed, the free energy landscape becomes comparable to the one from the equilibrium complex simulations only after the formation of those contacts as demonstrated in Figure S7.

Free Energy Landscape of the Nonspecific Binding.

We now turn our attention to the results based on the other nine spontaneous binding simulations. These are “unsuccessful” in the sense that the native intermolecular contacts barely formed (Q_{int} remains close to 0) in them; the folding of pKID (to misfolded states) can nevertheless occur as we demonstrate here. By combining all of those trajectories, we constructed the free energy landscape $f(n_{\text{contact}}, Q_{\text{pKID}})$ (surface colored yellow in Figure 5a). Here, we adopted the number of intermolecular heavy atom contacts (n_{contact}) as the binding coordinate, and the landscape $f(n_{\text{contact}}, Q_{\text{pKID}})$ was obtained by dividing the configuration space spanned by $0 \leq n_{\text{contact}} \leq 400$ and $0 \leq Q_{\text{pKID}} \leq 1$ into $50 \times 50 = 2500$ small areas and then computing the average f value for each small area; we also calculated the standard deviation $\sigma_f(n_{\text{contact}}, Q_{\text{pKID}})$ of f values sampled in each small area for an error estimate. We find that, also in these nonspecific bindings, the landscape becomes more progressively funneled as the binding proceeds; this can be inferred from the sections along Q_{pKID} taken at $n_{\text{contact}} = 0, 80$, and 160 (dashed lines in Figure 5a). The folding landscapes of pKID constructed from these sections (Figure 5b) visualize this finding. This indicates that the folding of pKID can occur on the KIX surface even in nonspecific complexes.

Indeed, in spite of an apparently dissimilar appearance of Figure 5a from Figure 3b, the free energy landscapes for the unsuccessful and successful trajectories are in fact comparable to each other. This is demonstrated in Figure 5c, in which the landscape from the successful trajectory redrawn with n_{contact} as the binding coordinate (surface colored brown) is superposed on top of the one from the unsuccessful trajectories (yellow). For a meaningful comparison, only the successful trajectory data up to the intermediate stage (<2250 ns) are used in the redrawing since the final specific complex stage, where the free energy f becomes considerably lower, is absent in the unsuccessful trajectories. Figure 5d shows that the two landscapes agree within statistical errors estimated by $\sigma_f(n_{\text{contact}}, Q_{\text{pKID}})$. Thus, topographic characteristics of the landscape associated with the initial folding of pKID are largely independent of the nature (native or non-native) of intermolecular contacts; i.e., they do not significantly depend on the angle of approach of pKID to KIX.

Projections of the two selected nonspecific binding trajectories onto the landscape are drawn with solid lines in Figure 5e,f; those of the other seven trajectories are presented in Figure S10. The final complex structures from these simulations are also displayed. We find that pKID is bound at various sites on the KIX surface in these nonspecific complexes. In three of the nine unsuccessful trajectories (two shown in Figure 5e,f and one in Figure S10), the folding of pKID into helical structures is observed. However, all of these “folded” structures are in fact misfolded ones, in which the angle between the α_A and α_B helices is ~ 60 – 70° . This corroborates the finding that the electrostatic interactions discussed in connection with Figure 4a, lacking in nonspecific complexes, are the crucial interactions in widening the α_A – α_B

angle toward $\sim 90^\circ$ to complete the coupled folding and binding of pKID.

DISCUSSION

A helical structure is, in general, not stable by itself: short helices taken out from stable globular proteins are found to be unstable when isolated.⁵⁰ Additional stabilizing interactions of the helical structure must, therefore, be present in stable globular proteins. For example, all of the three α helices in HP-35 are tightly in contact with its hydrophobic core.⁴⁴ On the other hand, intrinsically disordered proteins are generally characterized by a high content of polar and charged residues and by a low population of hydrophobic residues.^{51,52} As such, most of the residues in the pKID α_A helix are either polar or charged, and the bulky hydrophobic residues are somewhat localized in the pKID α_B helix region (Figure 2b); therefore, a ternary hydrophobic core that would stabilize the two-helix structure of the folded pKID cannot be formed. Such sequence characteristics of pKID clearly manifest themselves in the constructed free energy landscape: the landscape of pKID before binding is not steep enough to allow pKID to fold, and this is because pKID lacks additional stabilizing structural entities such as a hydrophobic core.

Upon binding with KIX, intermolecular interactions previously unavailable come into play. In particular, a hydrophobic core can now be formed intermolecularly at the binding interface (Figure 4a), which contributes to stabilizing the helical structure of pKID. It is therefore natural for the free energy landscape along the folding direction to become progressively more funneled as the binding proceeds. Thus, the free energy landscape constructed for the pKID–KIX system clearly shows that the folding of pKID occurs after pKID binds to KIX. This agrees with the experimental observation for this system.²⁷ We also find from the successful binding simulation that the binding is initiated by the anchoring of the disordered pKID α_B helix to the KIX surface, followed by the binding of the pKID α_A helix (Figure 2d). Such a sequence of events has been observed in previous experimental^{27,32} and computational^{33,37} studies. In this regard, we recognize an important role played by Tyr-658 of KIX in the early stages (Figure 4b). In fact, this residue is known to be the critical residue in pKID–KIX binding.²⁴ We also observe that the initial folding of pKID to misfolded states occurs even in nonspecific complexes at various sites on the KIX surface (Figure 5e,f and Figure S10). This is because the associated landscape topography is found to be largely independent of the nature (native or non-native) of intermolecular contacts (Figure 5c). This supports the conclusion of an earlier computational study indicating that non-native contacts can also help the coupled folding and binding of IDPs.⁵³

The key event, however, in completing the pKID–KIX coupled folding and binding elucidated in the present study is the correct orientation of the two α helices of pKID upon docking onto the surface of KIX, which should adopt a conformation where an angle of $\sim 90^\circ$ exists between them. The latter conformation is stabilized by electrostatic interactions (Figure 4a), which, in turn, require pKID to adopt the correct folded structure. Low-free-energy local structures and non-native intermolecular interactions formed in intermediate nonspecific complexes must be broken toward such a directed self-assembly, which leads to an increase in free energy. Indeed, we observe a free energy barrier in the pKID landscape that separates the misfolded intermediate and the

final folded states (Figure 3b,d). The presence of the metastable intermediate state in the pKID landscape is remarkable, as such an intermediate state is absent in the free energy landscape of most small proteins that exhibit the two-state folding behavior⁴⁷ (see, e.g., Figure 1a for HP-35). Our finding is corroborated by the experimental observation that a three-state model is necessary to account for the kinetics of the pKID–KIX coupled folding and binding.^{27,31}

Finally, caveats must be stated because, due to the high computational cost for carrying out spontaneous pKID–KIX binding simulations at a fully atomistic level, a number of analyses presented here are based on a single successful trajectory. For example, one could imagine that the native intermolecular contacts could have assembled in an order different from the one shown in Figure 4b, and even that the order observed in this work may not be the most common pathway. Furthermore, we cannot exclude a possibility that a barrier that separates the nonspecific and specific complex states may arise from configurational entropy as in protein folding (cf. Figure 1a,b). However, the landscape characteristics elucidated in the present work would not be significantly altered even if additional successful trajectories could be gained. Indeed, topographic characteristics (e.g., the landscape slope) of the configuration space are well-preserved even when underlying processes might be heterogeneous, as we argued in connection with Figure 1a and Figure S1 on the basis of a much longer simulation. We also observe that the landscape topography associated with the initial folding of pKID does not significantly depend on how pKID approaches KIX (Figure 5c), and that the landscape at the final specific complex state is consistent with the one obtained from the independent, equilibrium pKID–KIX complex simulations (Figure S7).

CONCLUSIONS

We demonstrate in the present work that molecular insights into the coupled folding and binding of an intrinsically disordered protein can be gained through an analysis of the free energy landscape. This is done by constructing the free energy landscape of the paradigmatic pKID–KIX system based on fully atomistic molecular dynamics simulations and the direct computation of the free energy that defines the landscape. We find that the free energy landscape of pKID before binding is not steep enough to allow pKID to fold by itself. This explains why pKID is disordered in a free state. Binding with KIX brings about considerable modifications to the landscape topography. In particular, the landscape along the pKID folding direction becomes progressively more funneled as the binding proceeds. This indicates that the folding of pKID occurs only after pKID binds to KIX. Additionally, we detect the presence of a metastable intermediate state in the landscape. This implies that the pKID–KIX coupled folding and binding involves at least three states. By linking these landscape characteristics to the structural details of the simulation trajectories, we provide a molecular mechanism by which disordered pKID folds upon binding to KIX. The present work further shows that the funneled landscape perspective, originally developed for ordered proteins, is enlightening also in advancing our understanding of intricate processes that involve intrinsically disordered proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.9b00200.

Free energy relations and equations, details on molecular dynamics simulations and structural analyses, the computational procedure of the solvation free energy, the construction of the free energy landscape, and additional data and figures including free energy landscapes, Q_{pKID} vs time, Q_{int} vs time, α -helical contents, angle between pKID α_A and α_B helices, free energy vs time, steps to construct the folding landscape, and an NMR complex structure (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: sihyun@sookmyung.ac.kr. Phone: +82 2 710 9410. Fax: +82 2 2077 7321.

ORCID

Sihyun Ham: 0000-0003-4717-7404

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Samsung Science and Technology Foundation under Project SSTF-BA1401-13.

REFERENCES

- (1) Lazaridis, T.; Karplus, M. Thermodynamics of protein folding: A microscopic view. *Biophys. Chem.* **2003**, *100*, 367–395.
- (2) Chong, S.-H.; Ham, S. Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis. *Acc. Chem. Res.* **2015**, *48*, 956–965.
- (3) Chong, S.-H.; Ham, S. Examining a thermodynamic order parameter of protein folding. *Sci. Rep.* **2018**, *8*, 7148.
- (4) Levinthal, C. How to fold graciously. *Mössbauer Spectroscopy in Biological Systems Proceedings* **1969**, *67*, 22–24.
- (5) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167–195.
- (6) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267*, 1619–1620.
- (7) Oliveberg, M.; Wolynes, P. G. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **2005**, *38*, 245–288.
- (8) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012**, *338*, 1042–1046.
- (9) Hartl, F. U.; Hayer-Hartl, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574–581.
- (10) Schug, A.; Onuchic, J. N. From protein folding to protein function and biomolecular binding by energy landscape theory. *Curr. Opin. Pharmacol.* **2010**, *10*, 709–714.
- (11) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. Predictive energy landscapes for protein–protein association. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19244–19249.
- (12) Nussinov, R.; Wolynes, P. G. A second molecular biology revolution? The energy landscapes of biomolecular function. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6321–6322.
- (13) Wang, J.; Oliveira, R. J.; Chu, X.; Whitford, P. C.; Chahine, J.; Han, W.; Wang, E.; Onuchic, J. N.; Leite, V. B. P. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15763–15768.

- (14) Chu, X.; Gan, L.; Wang, E.; Wang, J. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2342–E2351.
- (15) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. Challenges in protein-folding simulations. *Nat. Phys.* **2010**, *6*, 751–758.
- (16) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.
- (17) Ahmad, M.; Gu, W.; Geyer, T.; Helms, V. Adhesive water networks facilitate binding of protein interfaces. *Nat. Commun.* **2011**, *2*, 261.
- (18) Chong, S.-H.; Ham, S. Impact of chemical heterogeneity on protein self-assembly in water. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7636–7641.
- (19) Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346–3354.
- (20) Wright, P. E.; Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.
- (21) Uversky, V. N. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci.* **2013**, *22*, 693–724.
- (22) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5128.
- (23) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29.
- (24) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell* **1997**, *91*, 741–752.
- (25) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Dyson, H. J.; Wright, P. E. Conformational preferences in the Ser¹³³-kforms of the kinase inducible transactivation domain of CREB. *FEBS Lett.* **1998**, *430*, 317–322.
- (26) Goto, N. K.; Zor, T.; Martinez-Yamout, M.; Dyson, H. J.; Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). *J. Biol. Chem.* **2002**, *277*, 43168–43174.
- (27) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021–1025.
- (28) Brüschweiler, S.; Konrat, R.; Tollinger, M. Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core. *ACS Chem. Biol.* **2013**, *8*, 1600–1610.
- (29) Thakur, J. K.; Yadav, A.; Yadav, G. Molecular recognition by the KIX domain and its role in gene regulation. *Nucleic Acids Res.* **2014**, *42*, 2112–2125.
- (30) Shammass, S. L.; Travis, A. J.; Clarke, J. Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12055–12060.
- (31) Dahal, L.; Shammass, S. L.; Clarke, J. Phosphorylation of the IDP KID modulates affinity for KIX by increasing the lifetime of the complex. *Biophys. J.* **2017**, *113*, 2706–2712.
- (32) Dahal, L.; Kwan, T. O. C.; Shammass, S. L.; Clarke, J. pKID binds to KIX via an unstructured transition state with nonnative interactions. *Biophys. J.* **2017**, *113*, 2713–2722.
- (33) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* **2008**, *4*, No. e1000060.
- (34) Ganguly, D.; Chen, J. Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding. *J. Am. Chem. Soc.* **2009**, *131*, 5214–5223.
- (35) Chen, H.-F. Molecular dynamics simulation of phosphorylated KID post-translational modification. *PLoS One* **2009**, *4*, No. e6516.
- (36) Huang, Y.; Liu, Z. Nonnative interactions in coupled folding and binding processes of intrinsically disordered proteins. *PLoS One* **2010**, *5*, No. e15375.
- (37) Ganguly, D.; Chen, J. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 1251–1266.
- (38) Umezawa, K.; Ikebe, J.; Takano, M.; Nakamura, H.; Higo, J. Conformational ensembles of an intrinsically disordered protein pKID with and without a KIX domain in explicit solvent investigated by all-atom multicanonical molecular dynamics. *Biomolecules* **2012**, *2*, 104–121.
- (39) Kurcinski, M.; Kolinski, A.; Kmiecik, S. Mechanism of folding and binding of an intrinsically disordered protein as revealed by ab initio simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2224–2231.
- (40) Stanley, N.; Esteban-Martín, S.; De Fabritiis, G. Kinetic modulation of a disordered protein domain by phosphorylation. *Nat. Commun.* **2014**, *5*, 5272.
- (41) Bomblies, R.; Luitz, M. P.; Zacharias, M. Mechanism of pKID/KIX association studied by molecular dynamics free energy simulations. *J. Phys. Chem. B* **2016**, *120*, 8186–8192.
- (42) Ithuralde, R. E.; Roitberg, A. E.; Turjanski, A. G. Structured and unstructured binding of an intrinsically disordered protein as revealed by atomistic simulations. *J. Am. Chem. Soc.* **2016**, *138*, 8742–8751.
- (43) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17845–17850.
- (44) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* **1997**, *4*, 180–184.
- (45) Chong, S.-H.; Ham, S. Dissecting protein configurational entropy into conformational and vibrational contributions. *J. Phys. Chem. B* **2015**, *119*, 12623–12631.
- (46) Bicout, D. J.; Szabo, A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model. *Protein Sci.* **2000**, *9*, 452–465.
- (47) Karplus, M. Behind the folding funnel diagram. *Nat. Chem. Biol.* **2011**, *7*, 401–404.
- (48) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874–17879.
- (49) Ben-Naim, A. *Hydrophobic Interactions*; Plenum: New York, 1980.
- (50) Dill, K. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133–7155.
- (51) Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1090–1103.
- (52) Tompa, P. *Structure and Function of Intrinsically Disordered Proteins*; CRC Press: Boca Raton, 2010.
- (53) De Sancho, D.; Best, R. B. Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: Association of HIF1 α with CBP. *Mol. BioSyst.* **2012**, *8*, 256–267.